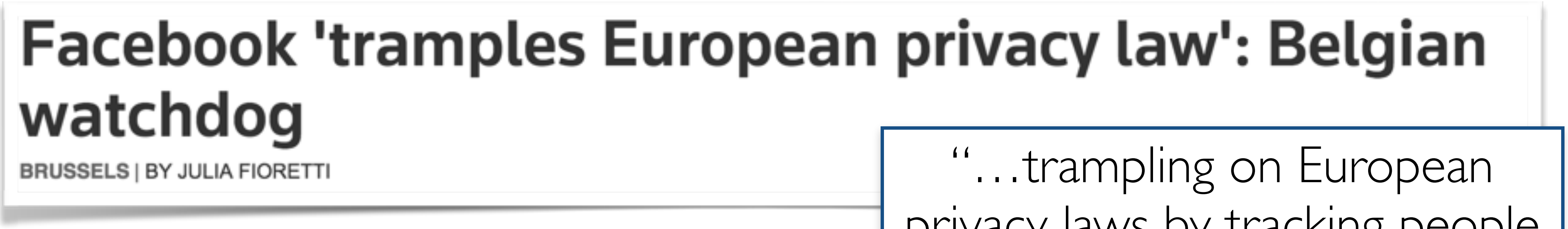
Variations in Tracking In Relation To Geographic Location

Nathaniel Fruchter
Hsin Miao
Scott Stevenson
Rebecca Balebako

W2SP 2015

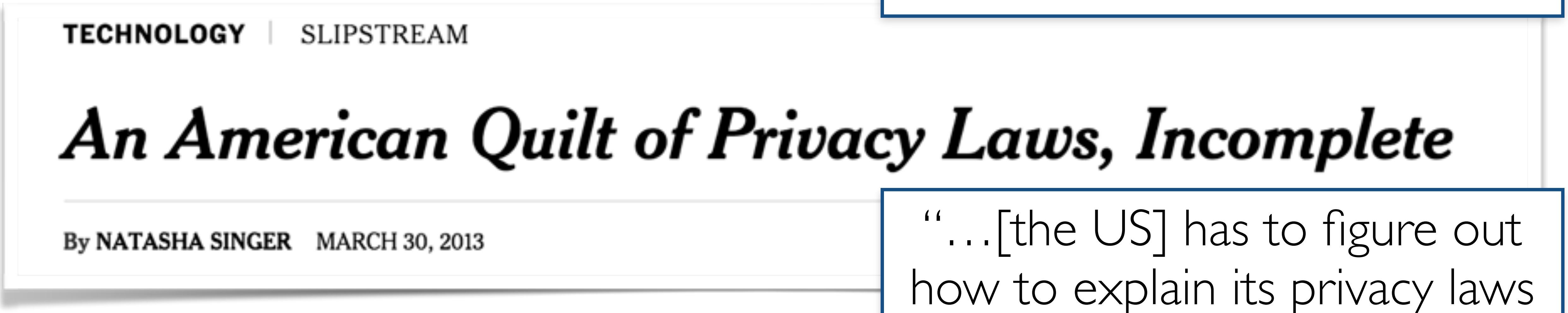
Carnegie Mellon University



Facebook 'tramples European privacy law': Belgian watchdog

BRUSSELS | BY JULIA FIORETTI

"…trampling on European privacy laws by tracking people online without their consent"

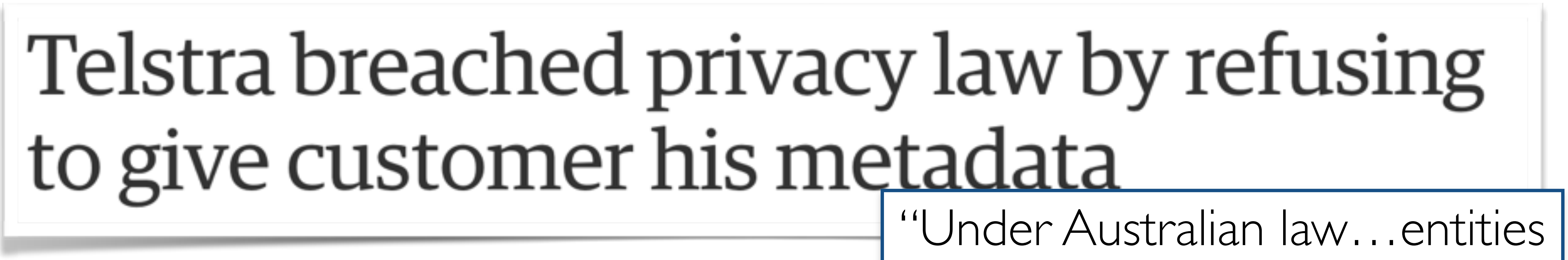


TECHNOLOGY | SLIPSTREAM

An American Quilt of Privacy Laws, Incomplete

By NATASHA SINGER MARCH 30, 2013

"…[the US] has to figure out how to explain its privacy laws on a global stage"



Telstra breached privacy law by refusing to give customer his metadata

"Under Australian law…entities must hand over 'personal information' they hold"

Governments have deemed privacy regulation necessary and feasible—it matters at the national and international level.

We need to think about how to evaluate its effectiveness.

The short version

- An empirical, automated method of measuring web tracking across countries
- Deployed in four countries representing three regulatory styles
- Significant differences found in amount of tracking
    - Where do these come from?

Coming up

- Privacy and legal regulation
- Measurement
    - Methods and heuristics
- Key observations
- Challenges and future work

Privacy and regulation

Privacy

- Third-party tracking of individuals has been recognized as a key issue when it comes to online privacy.

GHOSTERY

D. DISCONNECT

Lightbeam for Firefox

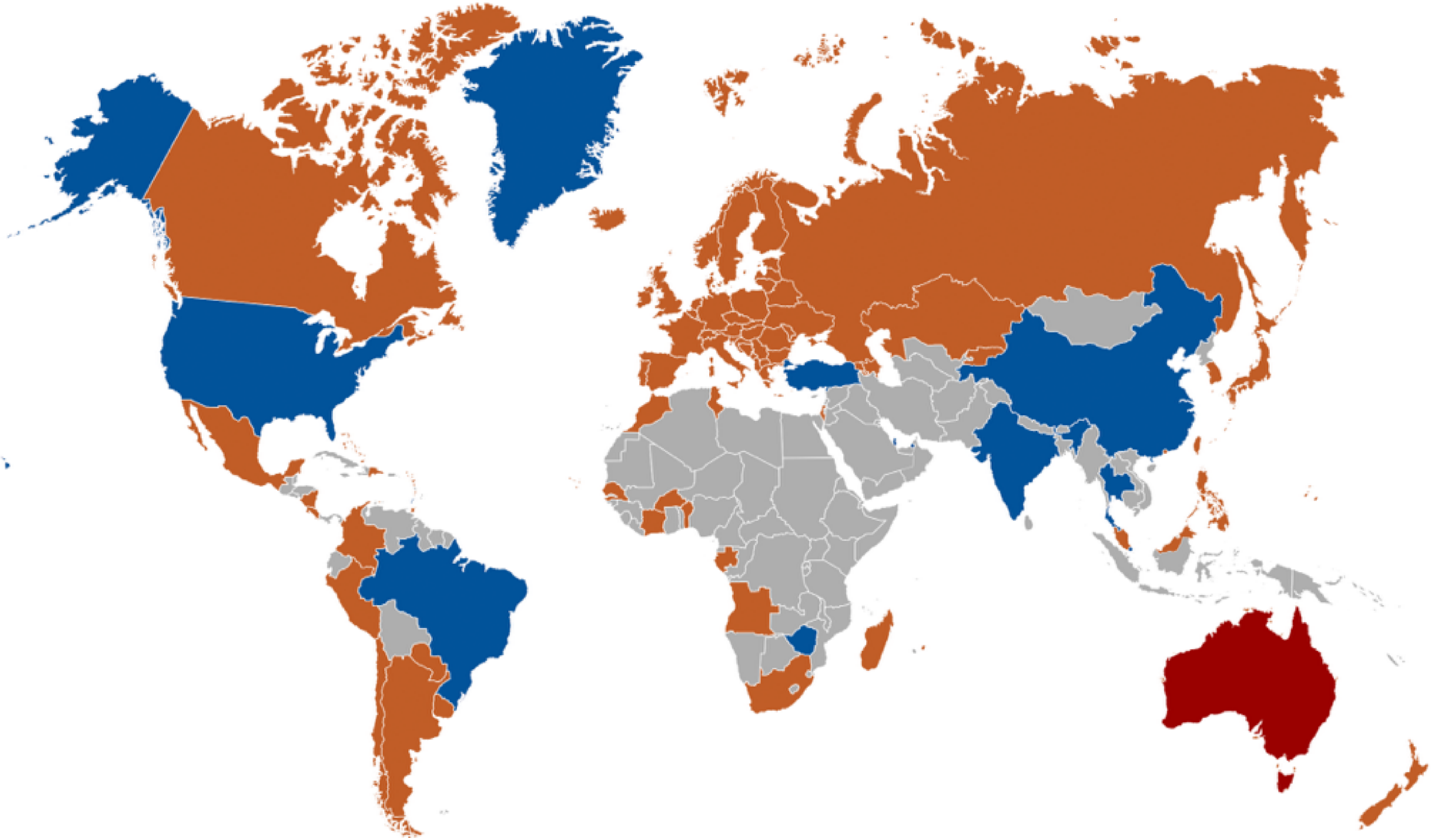
W3C

Tracking Preference Expression (DNT)

Privacy

- It’s **hard to define**.
- It’s an **incredibly relative concept**: culturally, personally, technologically…
- It’s an **incredibly dynamic concept** that changes along with many social and technological factors.

This doesn’t really make for the easiest landscape when it comes to regulatory action…



https://www.nymity.com/~/media/Nymity/Files/Privacy%20Maps/NYMITY_World_Map.ashx

Regulatory Regimes

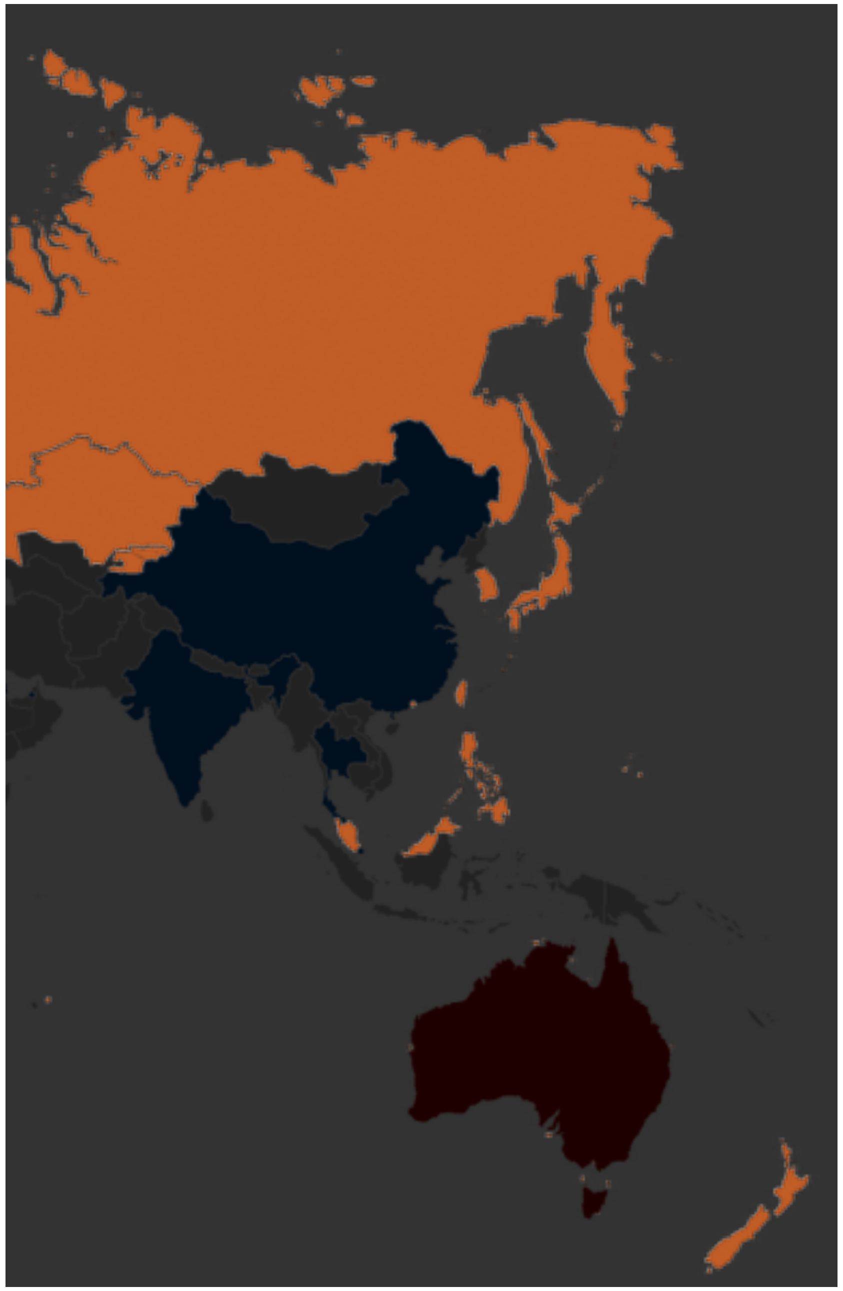
- Contrasting models of digital privacy regulation
- Different philosophies and methods!

Comprehensive

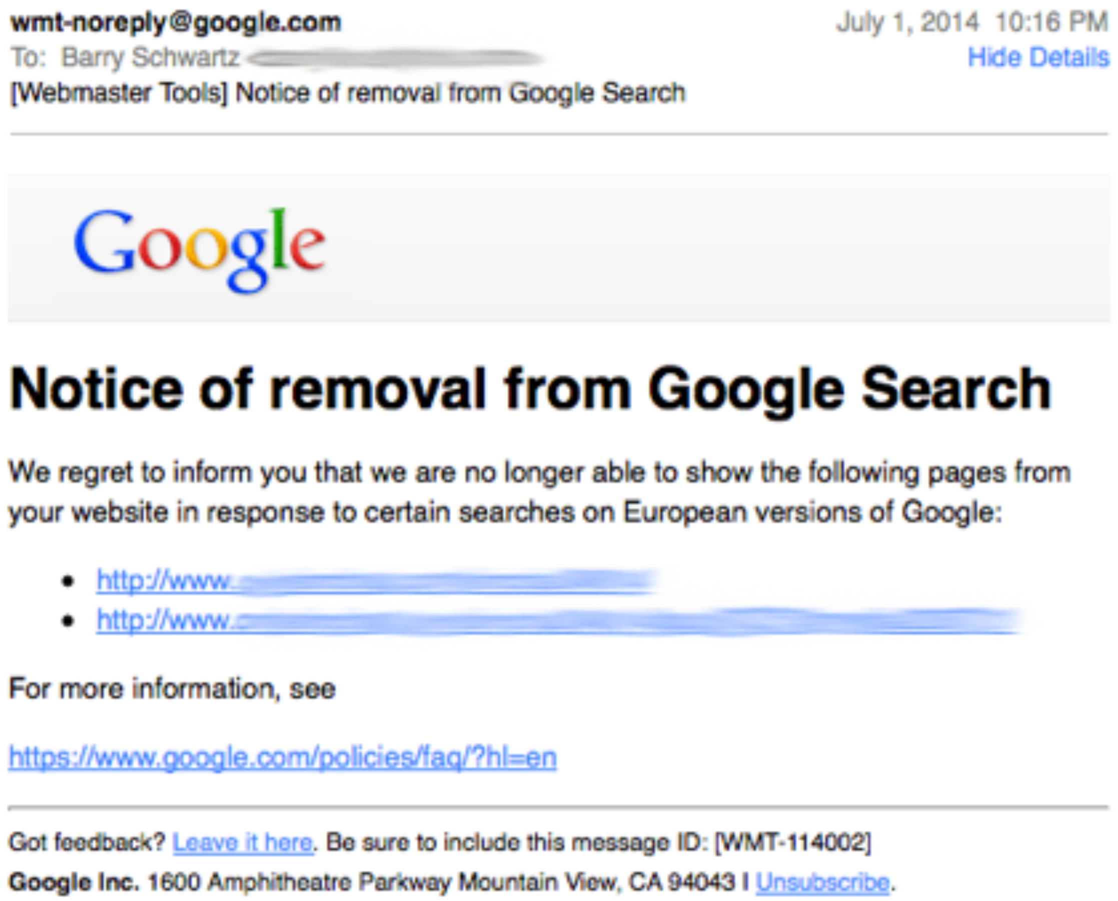
Regulatory Regimes

Comprehensive

- Privacy is a fundamental right.
- Legislated, top-down restrictions on collection, use, and disclosure.
- Enforced by dedicated regulatory bodies.



Office of the
Privacy Commissioner
of Canada



wmt-noreply@google.com July 1, 2014 10:16 PM
To: Barry Schwartz [redacted]

[Webmaster Tools] Notice of removal from Google Search

Notice of removal from Google Search

We regret to inform you that we are no longer able to show the following pages from your website in response to certain searches on European versions of Google:

- http://www.[redacted]
- http://www.[redacted]

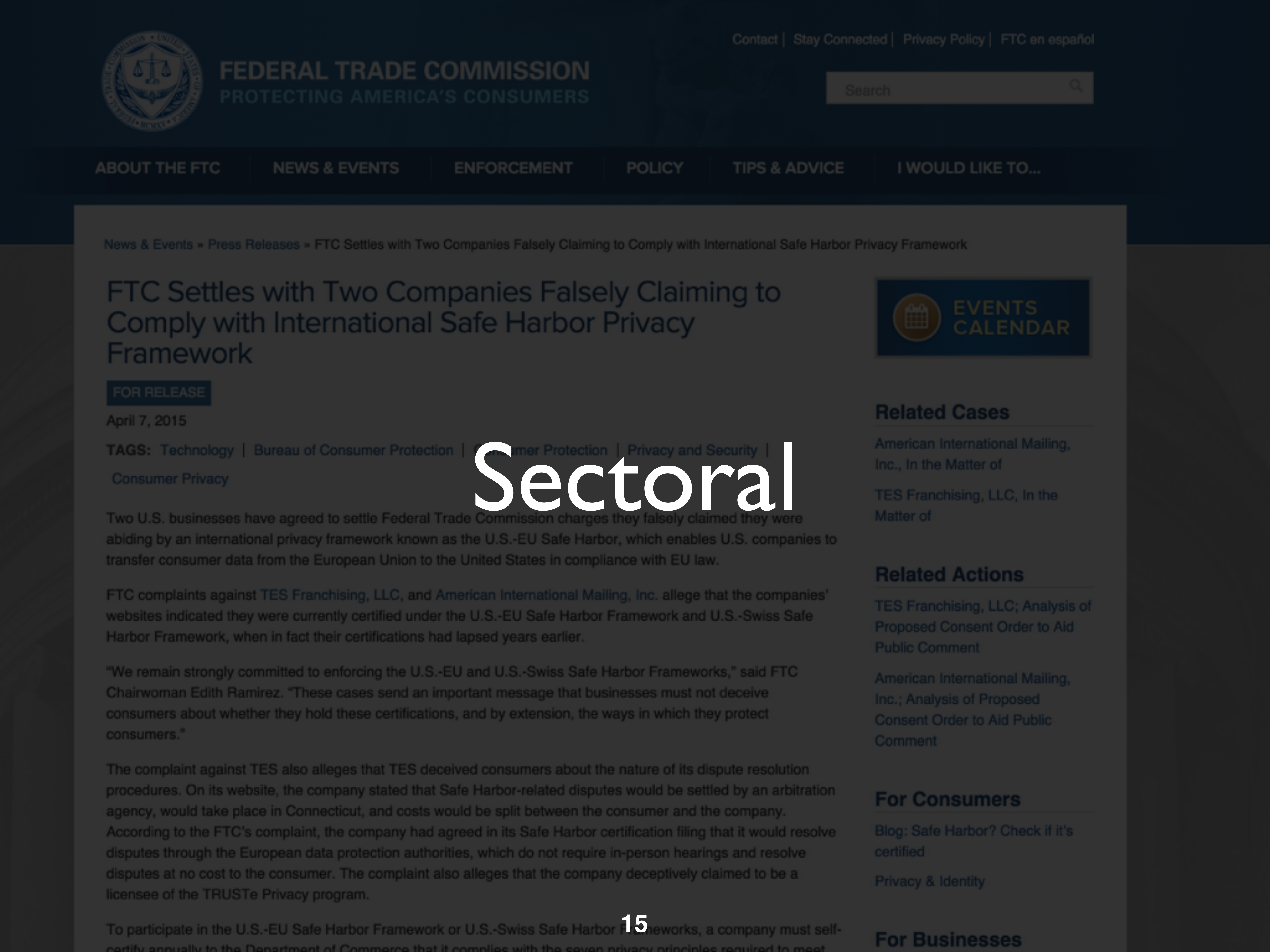
For more information, see

https://www.google.com/policies/faq/?hl=en

Got feedback? Leave it here. Be sure to include this message ID: [WMT-114002]
Google Inc. 1600 Amphitheatre Parkway Mountain View, CA 94043 I Unsubscribe.

European Data
Protection Supervisor

Sectoral



FEDERAL TRADE COMMISSION
PROTECTING AMERICA'S CONSUMERS
Contact | Stay Connected | Privacy Policy | FTC en español
Search
ABOUT THE FTC
NEWS & EVENTS
ENFORCEMENT
POLICY
TIPS & ADVICE
I WOULD LIKE TO...
News & Events » Press Releases » FTC Settles with Two Companies Falsely Claiming to Comply with International Safe Harbor Privacy Framework
EVENTS CALENDAR

FTC Settles with Two Companies Falsely Claiming to Comply with International Safe Harbor Privacy Framework

FOR RELEASE

April 7, 2015

TAGS: Technology | Bureau of Consumer Protection | Consumer Protection | Privacy and Security | Consumer Privacy

Two U.S. businesses have agreed to settle Federal Trade Commission charges they falsely claimed they were abiding by an international privacy framework known as the U.S.-EU Safe Harbor, which enables U.S. companies to transfer consumer data from the European Union to the United States in compliance with EU law.

FTC complaints against TES Franchising, LLC, and American International Mailing, Inc. allege that the companies' websites indicated they were currently certified under the U.S.-EU Safe Harbor Framework and U.S.-Swiss Safe Harbor Framework, when in fact their certifications had lapsed years earlier.

"We remain strongly committed to enforcing the U.S.-EU and U.S.-Swiss Safe Harbor Frameworks," said FTC Chairwoman Edith Ramirez. "These cases send an important message that businesses must not deceive consumers about whether they hold these certifications, and by extension, the ways in which they protect consumers."

The complaint against TES also alleges that TES deceived consumers about the nature of its dispute resolution procedures. On its website, the company stated that Safe Harbor-related disputes would be settled by an arbitration agency, would take place in Connecticut, and costs would be split between the consumer and the company. According to the FTC's complaint, the company had agreed in its Safe Harbor certification filing that it would resolve disputes through the European data protection authorities, which do not require in-person hearings and resolve disputes at no cost to the consumer. The complaint also alleges that the company deceptively claimed to be a licensee of the TRUSTe Privacy program.

To participate in the U.S.-EU Safe Harbor Framework or U.S.-Swiss Safe Harbor Frameworks, a company must self-certify annually to the Department of Commerce that it complies with the seven privacy principles required to meet

Related Cases

American International Mailing, Inc., In the Matter of

TES Franchising, LLC, In the Matter of

Related Actions

TES Franchising, LLC; Analysis of Proposed Consent Order to Aid Public Comment

American International Mailing, Inc.; Analysis of Proposed Consent Order to Aid Public Comment

For Consumers

Blog: Safe Harbor? Check if it's certified

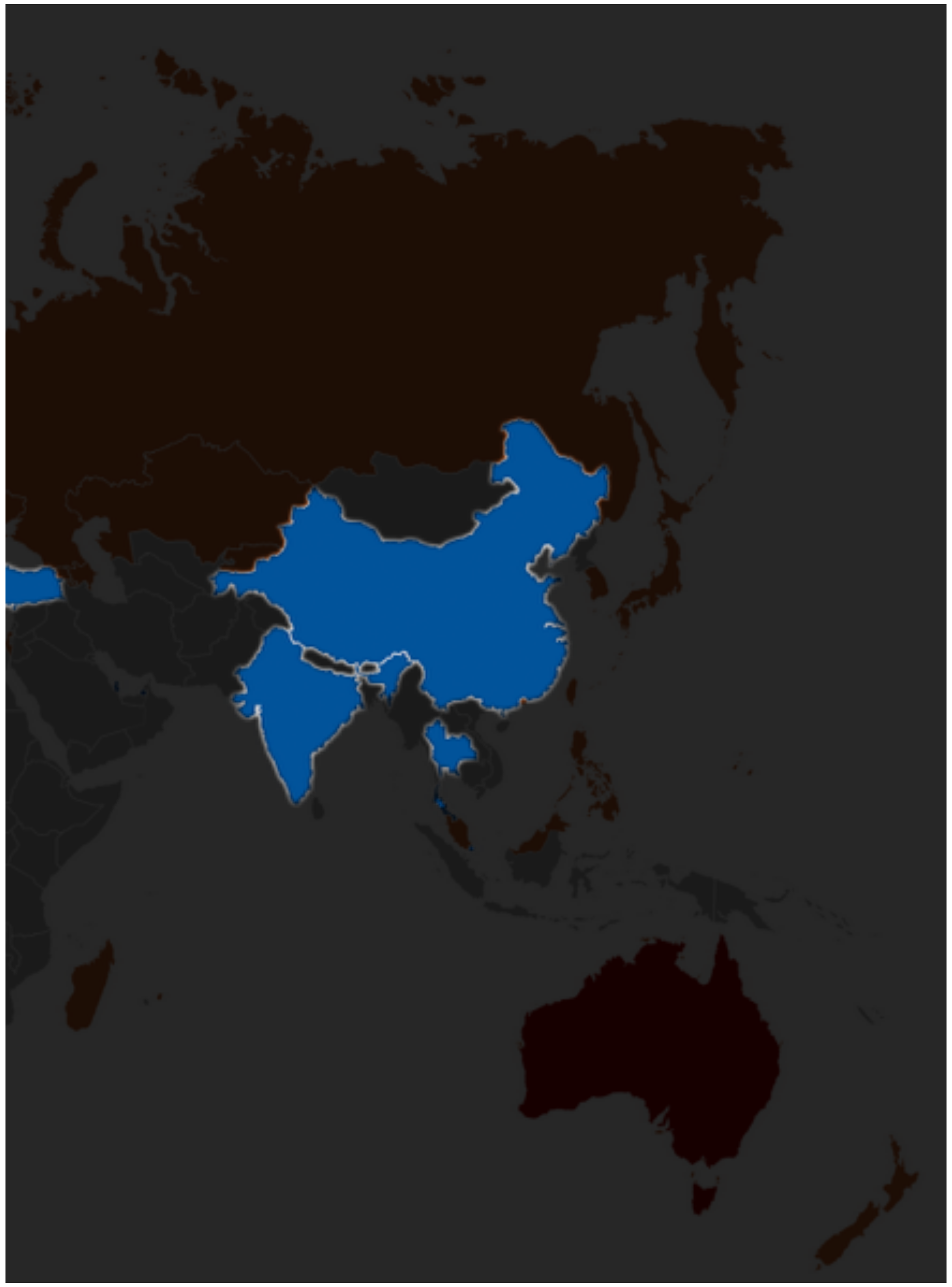
Privacy & Identity

For Businesses

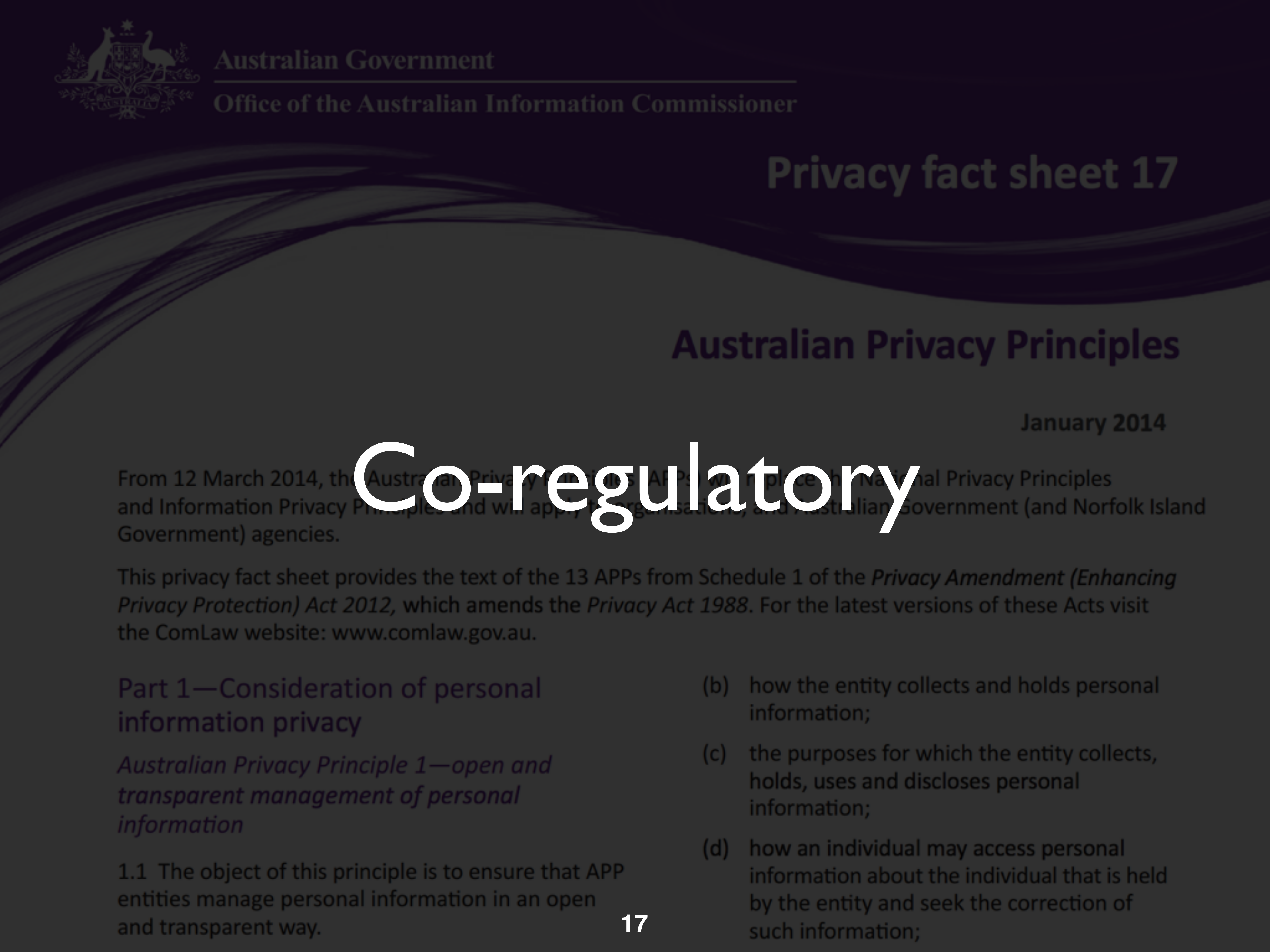
Regulatory Regimes

Sectoral

- Fewer fundamental protections.
- Privacy 'where it's needed': more of a patchwork.
    - Health, children, differences between US states.
- Emphasis on industry self-regulation and cooperation: "notice and choice"



Co-regulatory



Australian Government

Office of the Australian Information Commissioner

Privacy fact sheet 17

Australian Privacy Principles

January 2014

From 12 March 2014, the Australian Privacy Principles (APPs) will replace the National Privacy Principles and Information Privacy Principles and will apply to organisations, and Australian Government (and Norfolk Island Government) agencies.

This privacy fact sheet provides the text of the 13 APPs from Schedule 1 of the *Privacy Amendment (Enhancing Privacy Protection) Act 2012,* which amends the *Privacy Act 1988*. For the latest versions of these Acts visit the ComLaw website: www.comlaw.gov.au.

Part 1—Consideration of personal information privacy

Australian Privacy Principle 1—open and transparent management of personal information

1.1 The object of this principle is to ensure that APP entities manage personal information in an open and transparent way.

(b) how the entity collects and holds personal information;

(c) the purposes for which the entity collects, holds, uses and discloses personal information;

(d) how an individual may access personal information about the individual that is held by the entity and seek the correction of such information;

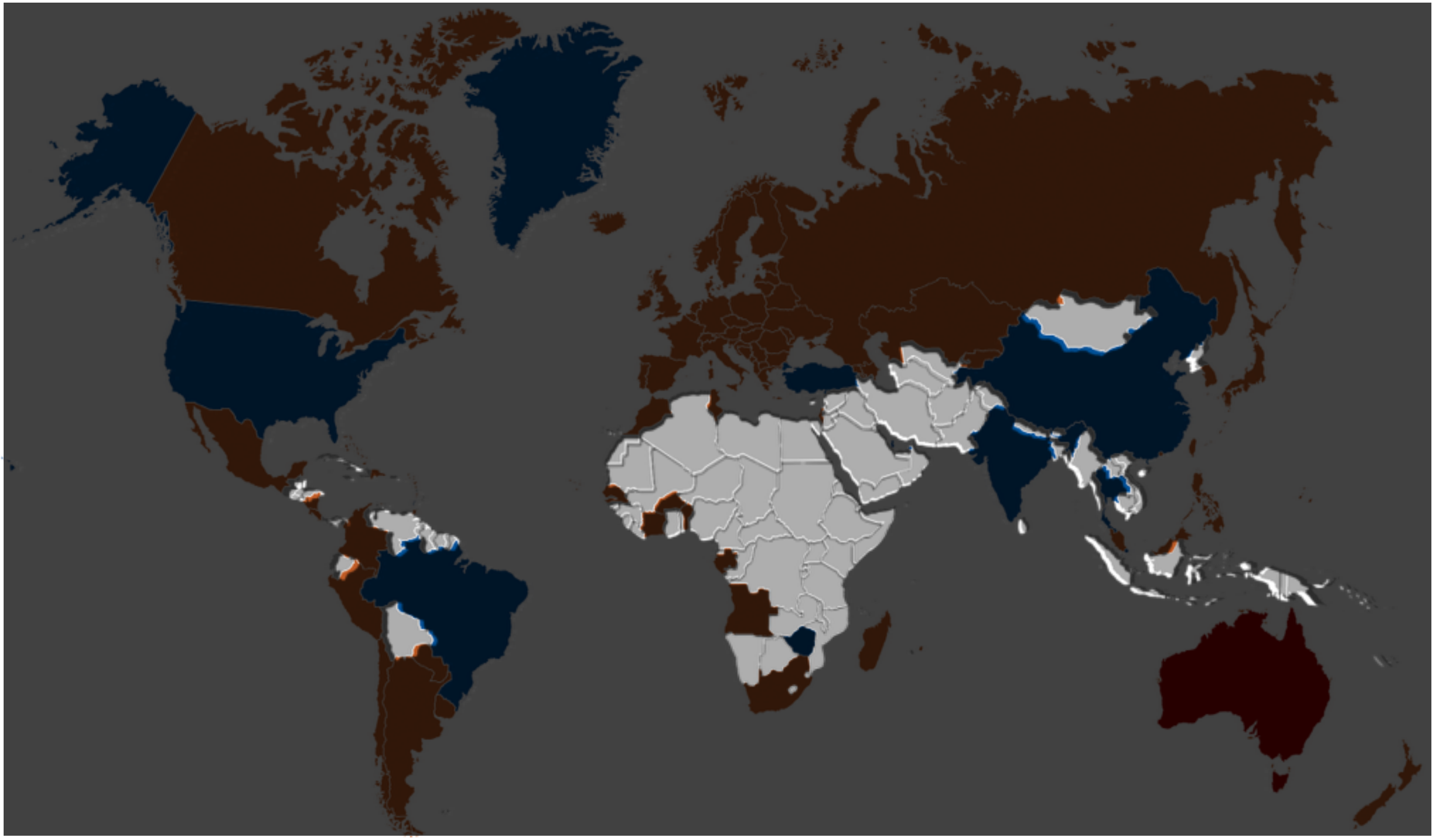
Regulatory Regimes

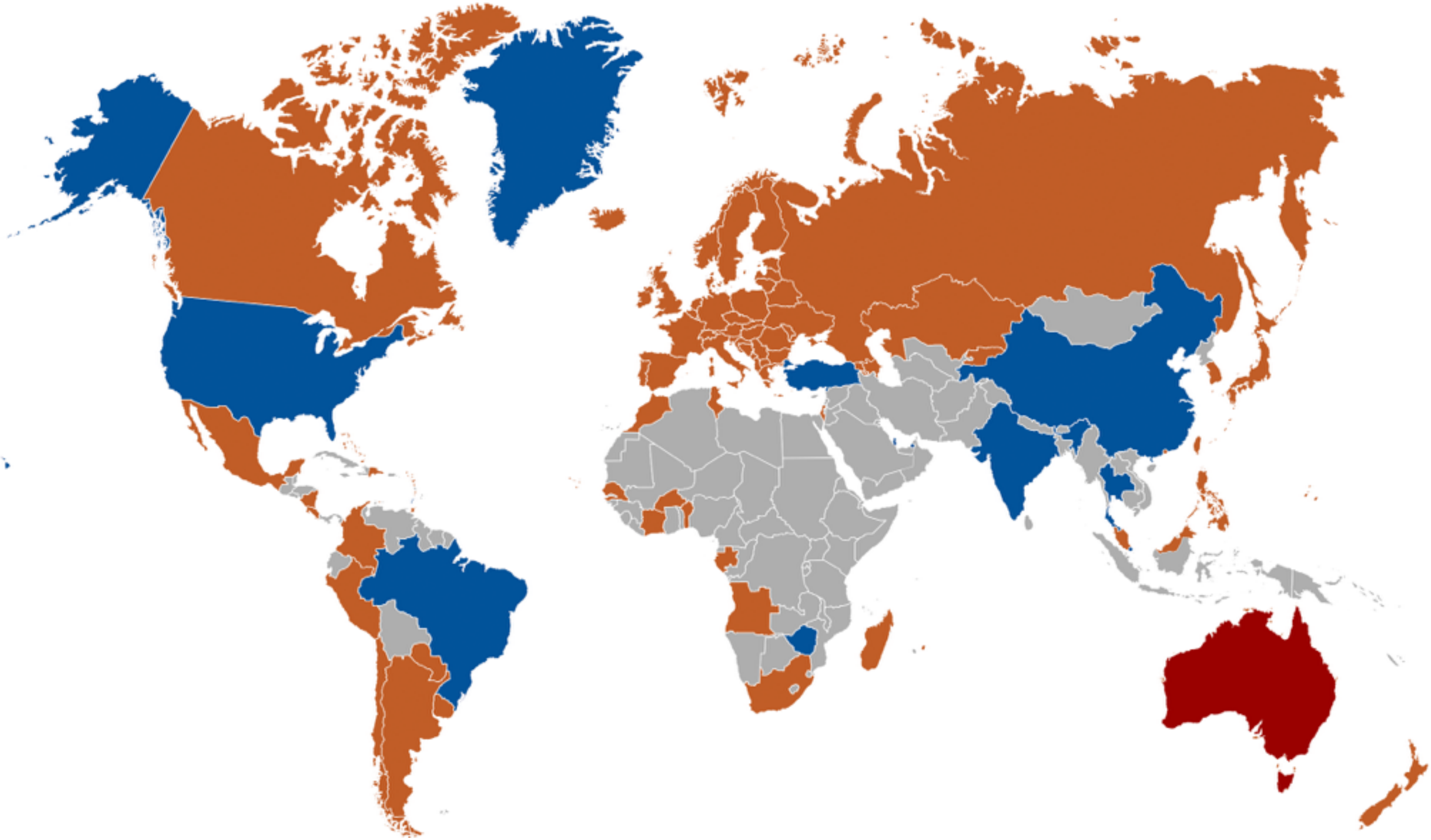
Co-regulatory

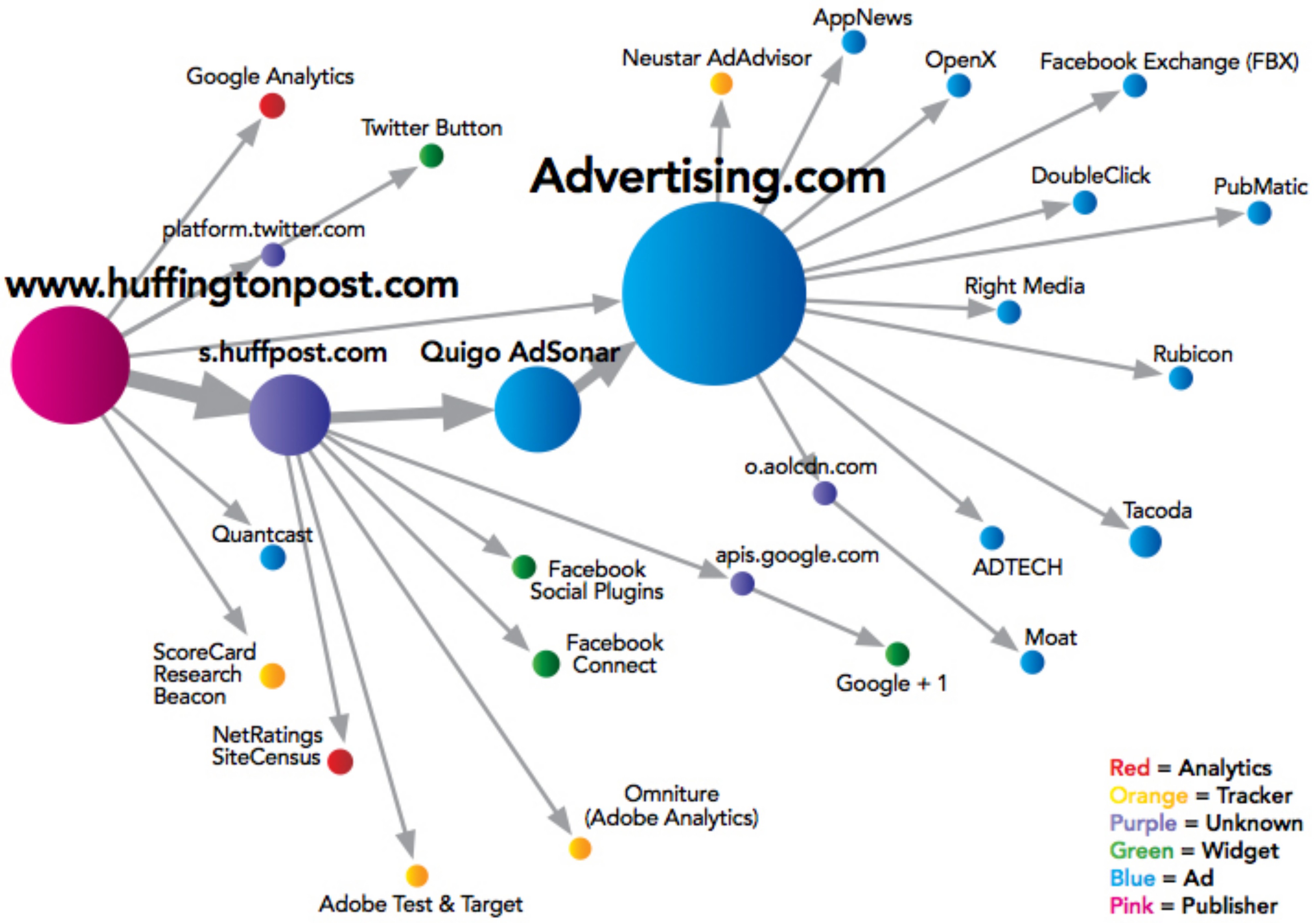
- Reliance on industry self-regulation with a government "backstop"
- Industry bound to create enforceable codes
- Most notably in Australia (but changing)

Regulatory Regimes

None or other







Evidon / Ghostery Enterprise, 2014

Do these regulatory (and geographic) differences lead to any quantifiable impact in web privacy and tracking?

Do these regulatory (and geographic) differences lead to any quantifiable impact in web privacy and tracking?

What is driving these differences?

Web measurement methods

Web measurement

- Measuring what the user (and their browser) actually sees and receives

- Assessing and quantifying what happens "in the wild" in a variety of situations

Our approach
Overview

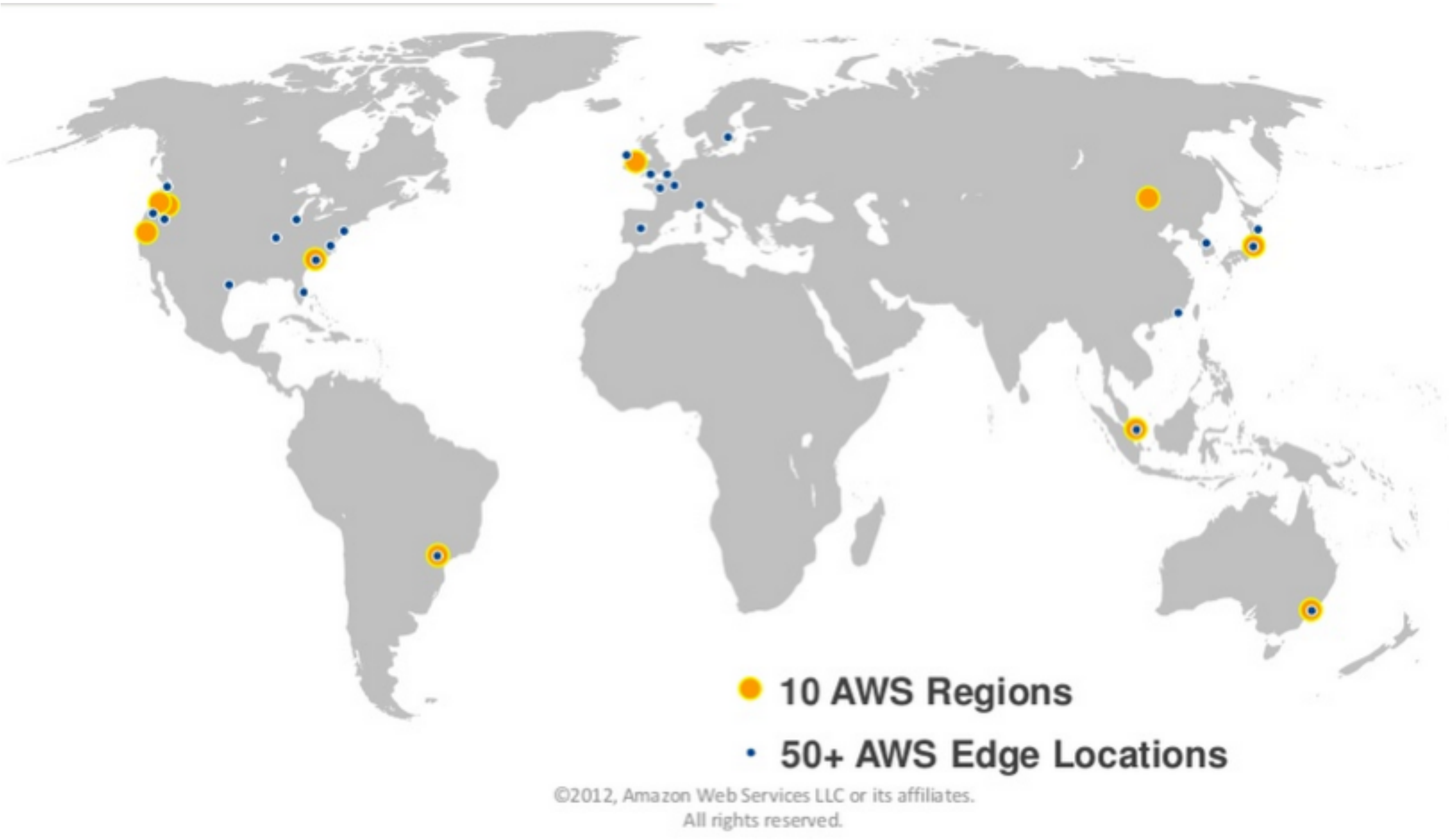
- Standardized
  - Python + OpenWPM library
- Reproducible
  - Open source, scripted
- Empirical
  - Controlled, automated, no humans
- Realistic*
  - Flash, JavaScript, Firefox engine

Our approach
Network infrastructure

- How do you source a network endpoint in different countries without introducing extra measurement confounds?

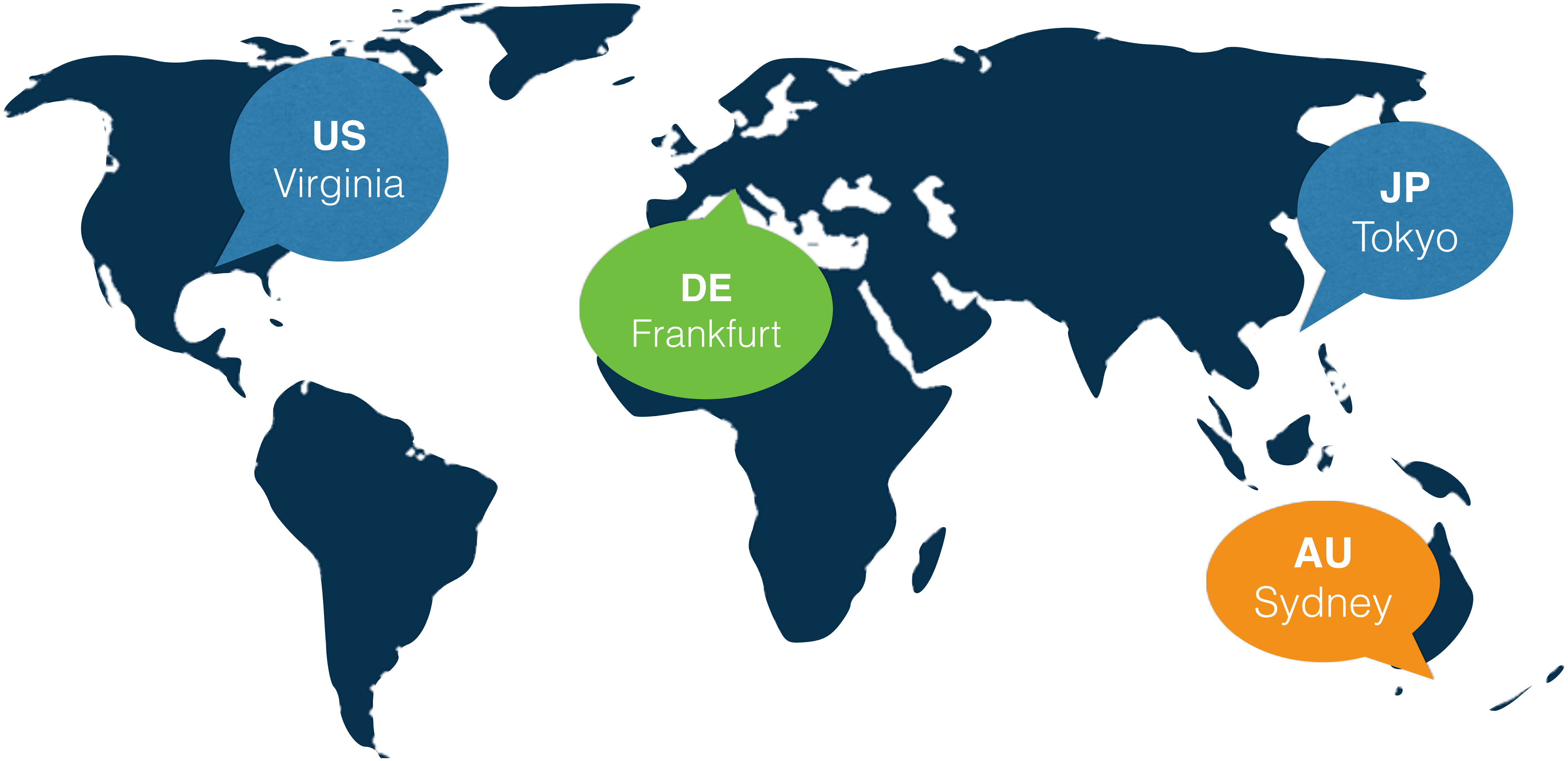
Our approach

Network infrastructure



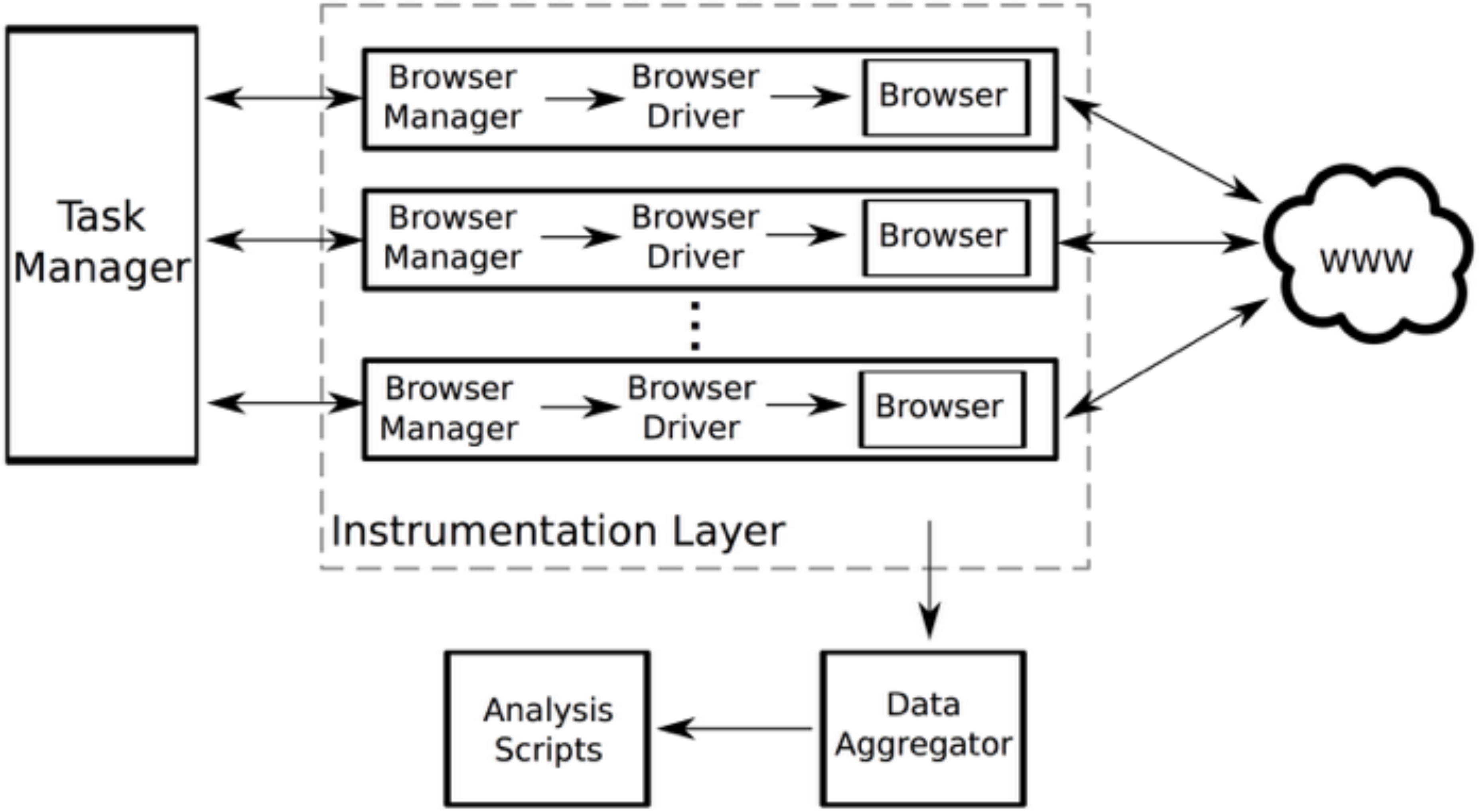
Our approach

Network infrastructure



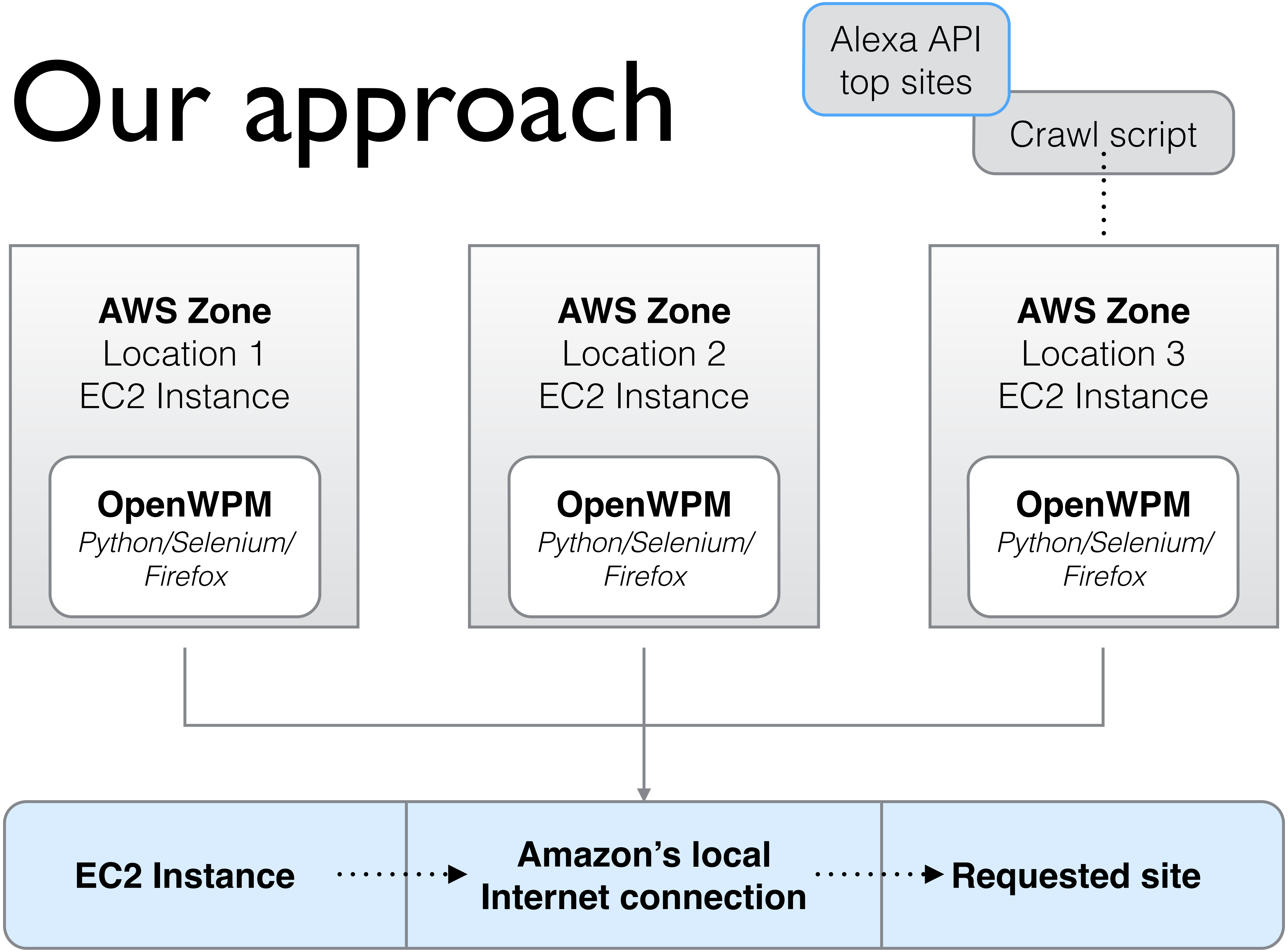
Sectoral | Comprehensive | Co-regulatory

OpenWPM 0.2.1
(Engelhardt et al, 2014)



http://randomwalker.info/publications/WebPrivacyMeasurement.pdf

Our approach



Alexa API top sites

Crawl script

AWS Zone
Location 1
EC2 Instance

OpenWPM
Python/Selenium/ Firefox

AWS Zone
Location 2
EC2 Instance

OpenWPM
Python/Selenium/ Firefox

AWS Zone
Location 3
EC2 Instance

OpenWPM
Python/Selenium/ Firefox

EC2 Instance → **Amazon's local Internet connection** → **Requested site**

Our approach
Heuristics

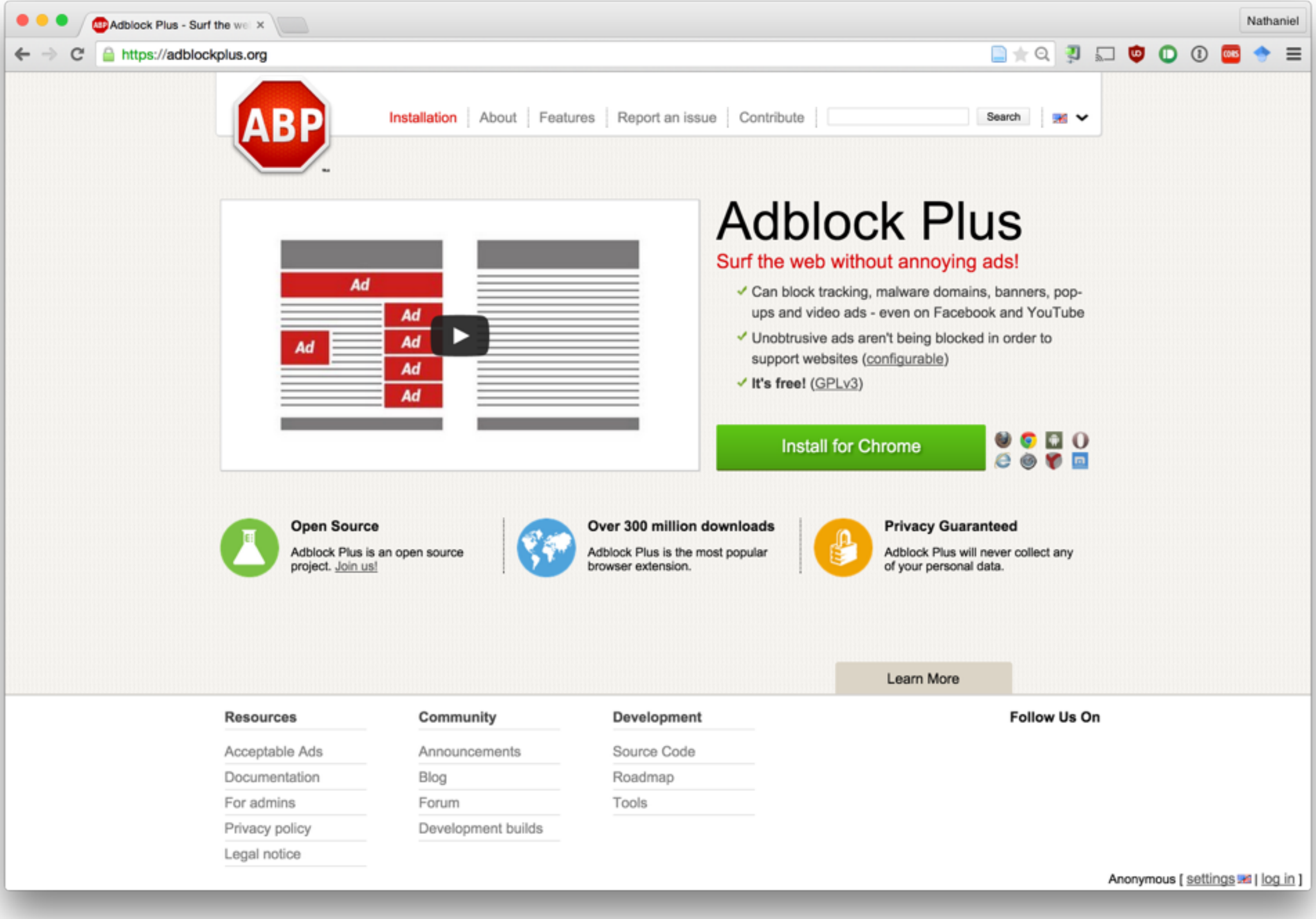
- Measure: **third-party HTTP requests + cookies**
  - First-party requests have been exempted from definition of tracking/advertising (Do Not Track specification*)
  - Rough metric, but can be representative

*McDonald and Peha (2011),
“Track Gap: Policy Implications of User Expectations for the `Do Not Track’ Internet Privacy Feature”

Our approach

Heuristics

- Approach A: **simple count**
- Approach B: match against a large **database of web assets** generally agreed upon as tracking



Adblock Plus - Surf the we ×
Nathaniel
https://adblockplus.org
ABP
Installation | About | Features | Report an issue | Contribute | Search
Ad
Ad
Ad
Ad
Ad
Ad

Adblock Plus

Surf the web without annoying ads!

- ✓ Can block tracking, malware domains, banners, pop-ups and video ads - even on Facebook and YouTube
- ✓ Unobtrusive ads aren't being blocked in order to support websites (configurable)
- ✓ **It's free!** (GPLv3)

Install for Chrome

Open Source

Adblock Plus is an open source project. Join us!

Over 300 million downloads

Adblock Plus is the most popular browser extension.

Privacy Guaranteed

Adblock Plus will never collect any of your personal data.

Learn More

Resources

- Acceptable Ads
- Documentation
- For admins
- Privacy policy
- Legal notice

Community

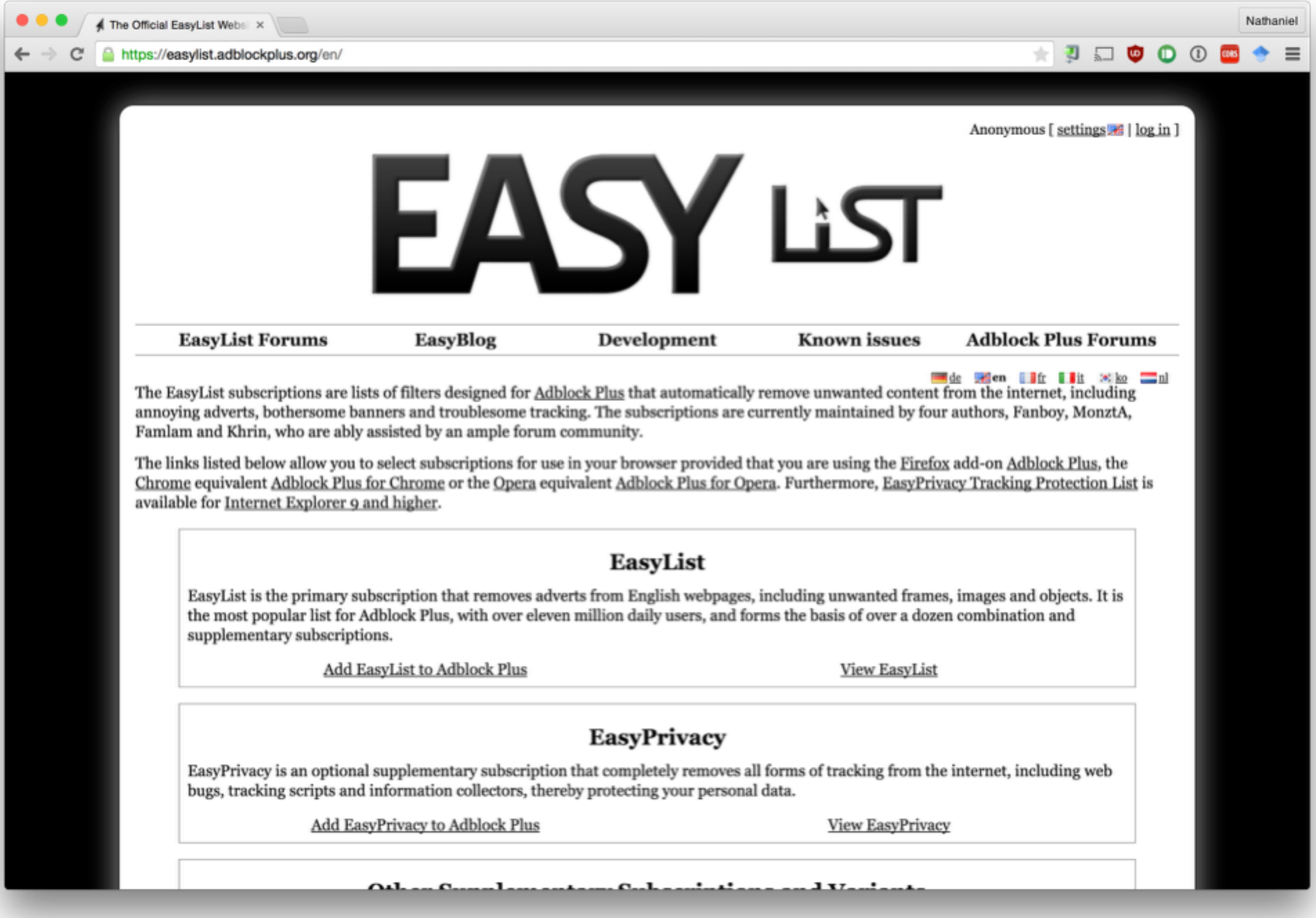
- Announcements
- Blog
- Forum
- Development builds

Development

- Source Code
- Roadmap
- Tools

Follow Us On

Anonymous [settings | log in]


The Official EasyList Webs
Nathaniel
https://easylist.adblockplus.org/en/

Anonymous [settings | log in]

EASY LIST

EasyList Forums | EasyBlog | Development | Known issues | Adblock Plus Forums

de en fr it ko nl

The EasyList subscriptions are lists of filters designed for Adblock Plus that automatically remove unwanted content from the internet, including annoying adverts, bothersome banners and troublesome tracking. The subscriptions are currently maintained by four authors, Fanboy, MonztA, Famlam and Khrin, who are ably assisted by an ample forum community.

The links listed below allow you to select subscriptions for use in your browser provided that you are using the Firefox add-on Adblock Plus, the Chrome equivalent Adblock Plus for Chrome or the Opera equivalent Adblock Plus for Opera. Furthermore, EasyPrivacy Tracking Protection List is available for Internet Explorer 9 and higher.

EasyList

EasyList is the primary subscription that removes adverts from English webpages, including unwanted frames, images and objects. It is the most popular list for Adblock Plus, with over eleven million daily users, and forms the basis of over a dozen combination and supplementary subscriptions.

Add EasyList to Adblock Plus | View EasyList

EasyPrivacy

EasyPrivacy is an optional supplementary subscription that completely removes all forms of tracking from the internet, including web bugs, tracking scripts and information collectors, thereby protecting your personal data.

Add EasyPrivacy to Adblock Plus | View EasyPrivacy

Other Supplementary Subscriptions and Variants

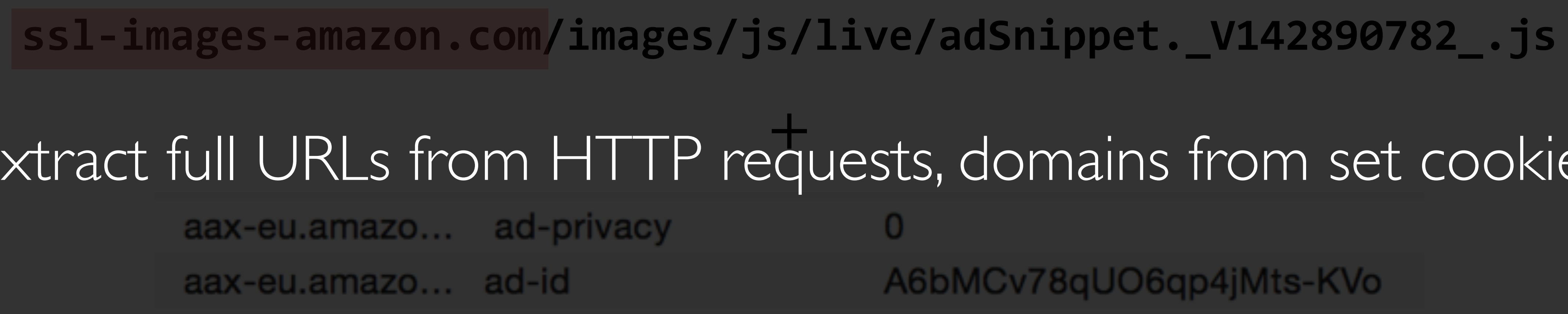
Our approach
Heuristics

- Approach B: parse and match against **open-source ad blocking rulesets**
- We chose EasyList, the most commonly used and distributed AdBlock list
  - EasyList Ads and EasyPrivacy list
  - Over 50,000 regex-based rules
- *adblockparser* Python module*

* *https://github.com/scrapinghub/adblockparser*

Our approach

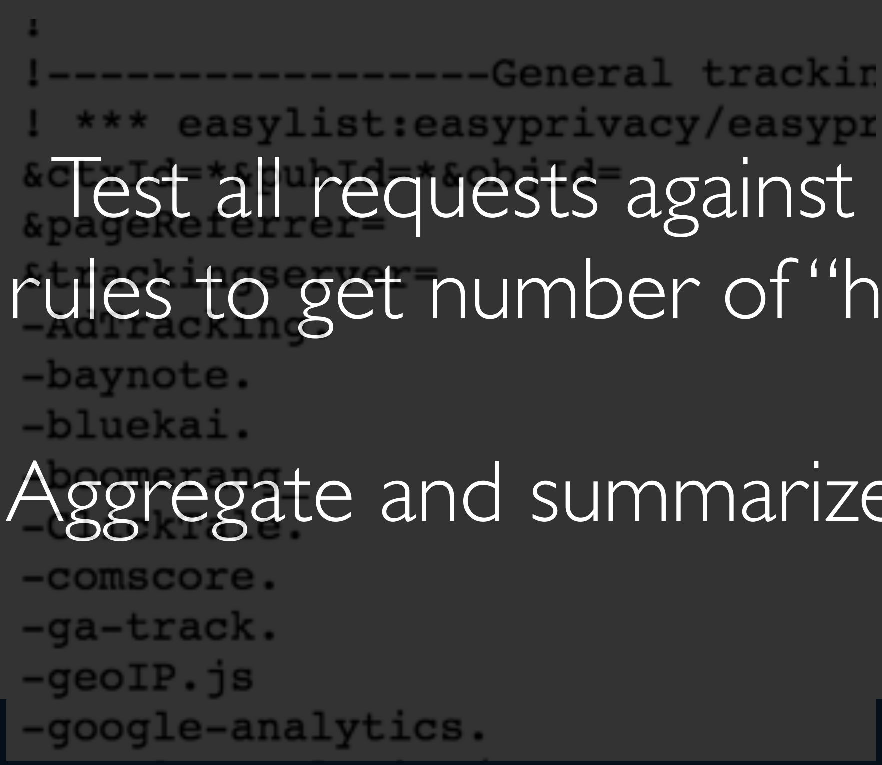
Analysis



```
ssl-images-amazon.com/images/js/live/adSnippet._V142890782_.js
```

aax-eu.amazo...	ad-privacy	0
aax-eu.amazo...	ad-id	A6bMCv78qUO6qp4jMts-KVo

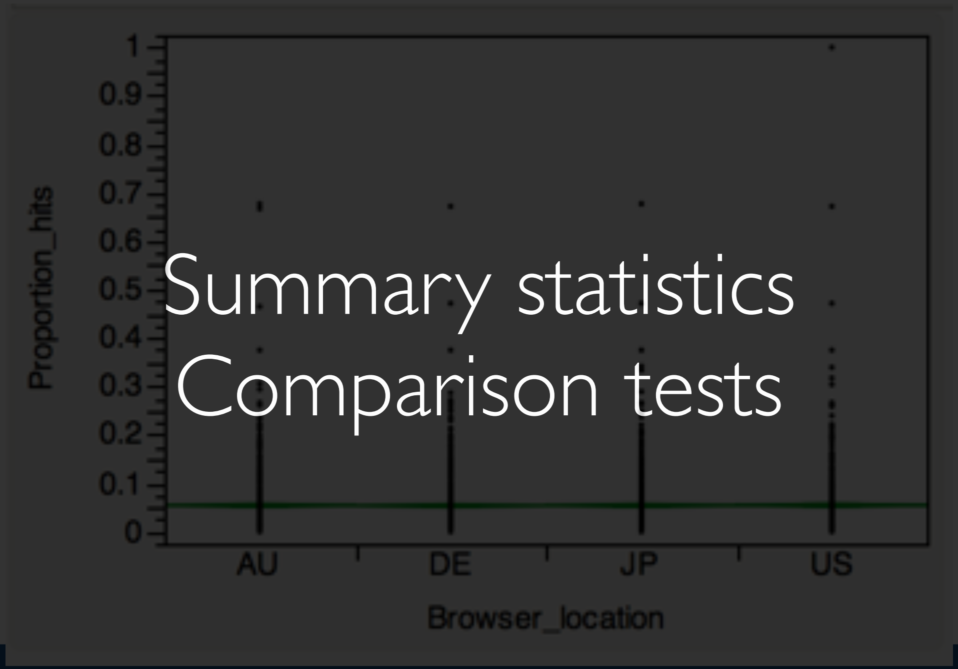
Extract full URLs from HTTP requests, domains from set cookies



```
!
!-----------------General trackin
! *** easylist:easyprivacy/easypr
&ctxId=*&pubId=*&objId=
&pageReferrer=
&trackingserver=
-AdTracking.
-baynote.
-bluekai.
-boomerang.
-Clickrate.
-comscore.
-ga-track.
-geoIP.js
-google-analytics.
```

Test all requests against
all rules to get number of "hits"

Aggregate and summarize



Summary statistics
Comparison tests

Key observations

Third-party requests/cookies

- Rank test against totals and ratios

	Tracking Indicator **Requests**	Tracking Indicator **Cookies**
US	1	1
AU	2	-
DE	-	-
JP	3	-

- Dash indicates a tie

Third-party requests/cookies

- The United States has significantly more activity across both metrics
- Interesting differences across countries
    - Caveat: sample representativeness

Ad blocking rules

Country-level results

Country	*Average* **requests/page**	*Average* **hits/page**	*Normalized* **% hits**
US	120.6	9.3	**8%**
AU	99.2	6.8	**6%**
DE	121.0	5.7	**5%**
JP	103.2	4.1	**5%**

Ad blocking rules

Country-level results

Country A	Country B	Compare A to B	
US	JP	2.8 to 4.0%	more
US	DE	1.8 to 3.1%	
US	AU	0.1% to 1.4%	
JP	DE	0.2 to 1.3%	less
DE	AU	0.9 to 2.1%	

Ad blocking rules

Results

- Significant differences between all pairs of countries
  - United States: more activity in all cases
    - 0.1% compared to Australia
    - 4% compared to Japan
  - 4% x ~100 average requests = **4+ tracking elements**
- Side note: more trackers than ads

Ad blocking rules

Origin-dependent activity

- Does tracking activity change depending on the origin of the user *or* the origin of the website?
- How much do we need to control for geographic factors?
- Synchronized crawl of top 500 global websites (same sites, different countries)
- No significant differences!

Limitations
and further work

The policy lifecycle

- **Development**: Recognize and diagnose the problem, identify and evaluate options
- **"In the wild"**: Implement, enforce, **monitor** (the hard part)

Limitations

Looking at privacy regulation

- Is our idea of what to expect from regulatory models correct?
- Is the (narrow) viewpoint that we tested where we would see the effect?

Limitations

Looking at privacy regulation

- US vs. Japan: sectoral vs. sectoral
  - Why does the US have more tracking?
  - Cultural practices, business norms, "Internet ecosystem", what's popular….

Limitations

Web measurement

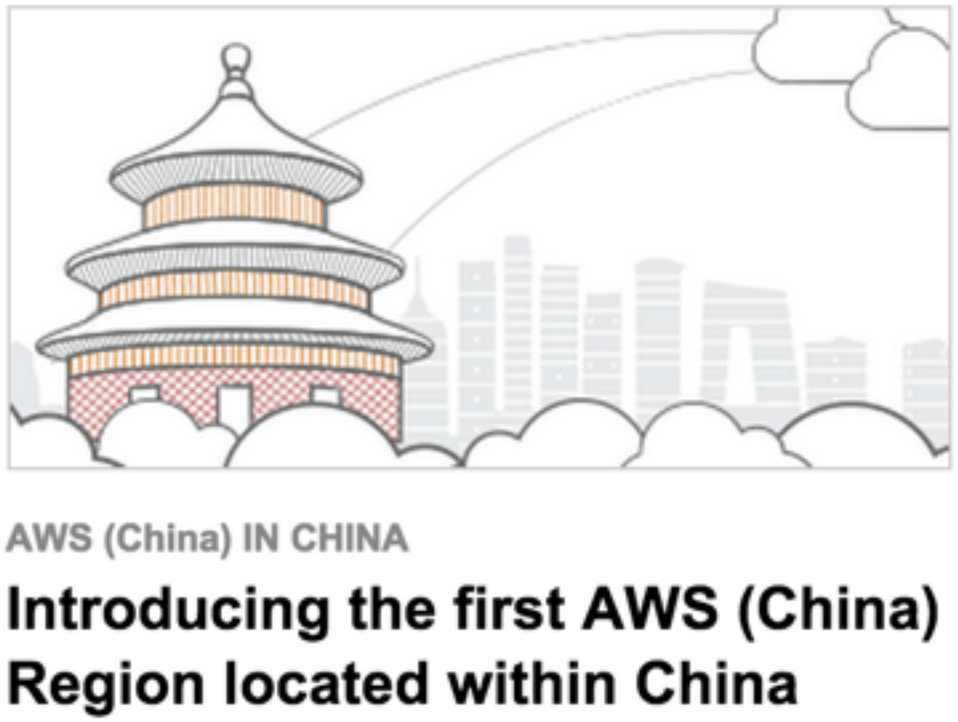
- What if we had a different Internet landscape?
  - China and other interesting locations

Bai du 百度
www.baidu.com
腾讯网
QQ.com
淘宝网
Taobao.com
新浪微博
weibo.com
youku 优酷
.com

Limitations

Web measurement

- More representative sample of networks!
- Amazon AWS has a limited number of availability zones
  - Promising developments?



AWS (China) IN CHINA

Introducing the first AWS (China) Region located within China

Limitations

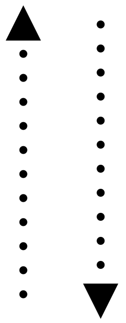
Web measurement

- Web activity is deterministic
  - Controls: automated "clean slate" for measurement
- Is first-party still a relevant distinction?
  - Inter-session, inter-device, and more pervasive forms of tracking

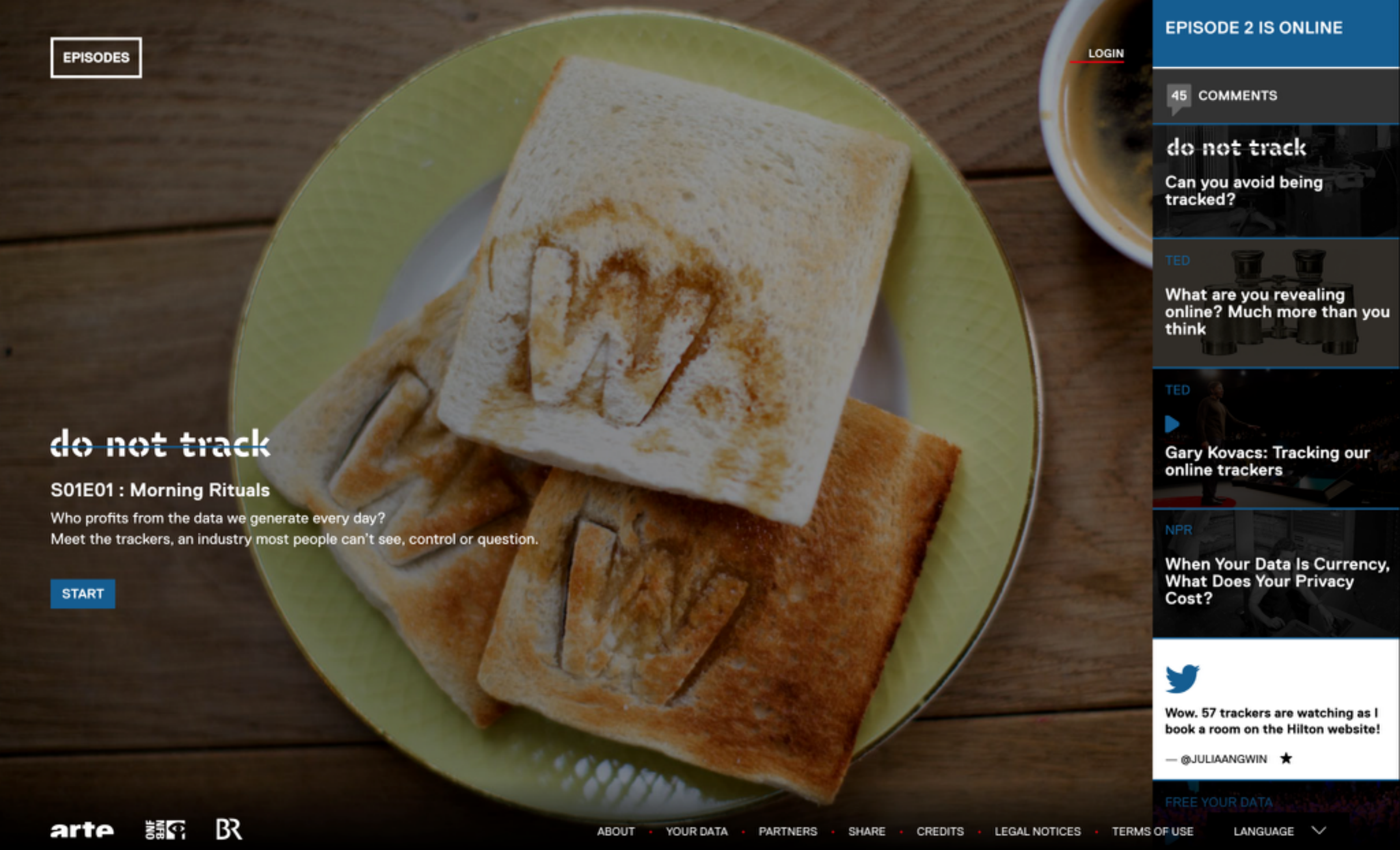
Next steps

- Limited sampling base (more connections needed!)
- Deeper exploration of differences:
  - Within regulatory models, cultural and business practices…
- You can always use more controls.
- Replication!

We need to think about how to evaluate effectiveness.



How effective are these models at providing what we want and expect?



EPISODES

LOGIN

EPISODE 2 IS ONLINE

45 COMMENTS

do not track

Can you avoid being tracked?

TED

What are you revealing online? Much more than you think

TED

Gary Kovacs: Tracking our online trackers

NPR

When Your Data Is Currency, What Does Your Privacy Cost?

Wow. 57 trackers are watching as I book a room on the Hilton website!

— @JULIAANGWIN ★

FREE YOUR DATA

do not track

S01E01 : Morning Rituals

Who profits from the data we generate every day?
Meet the trackers, an industry most people can't see, control or question.

START

arte

ABOUT · YOUR DATA · PARTNERS · SHARE · CREDITS · LEGAL NOTICES · TERMS OF USE   LANGUAGE

https://donottrack-doc.com *(April 2015)*

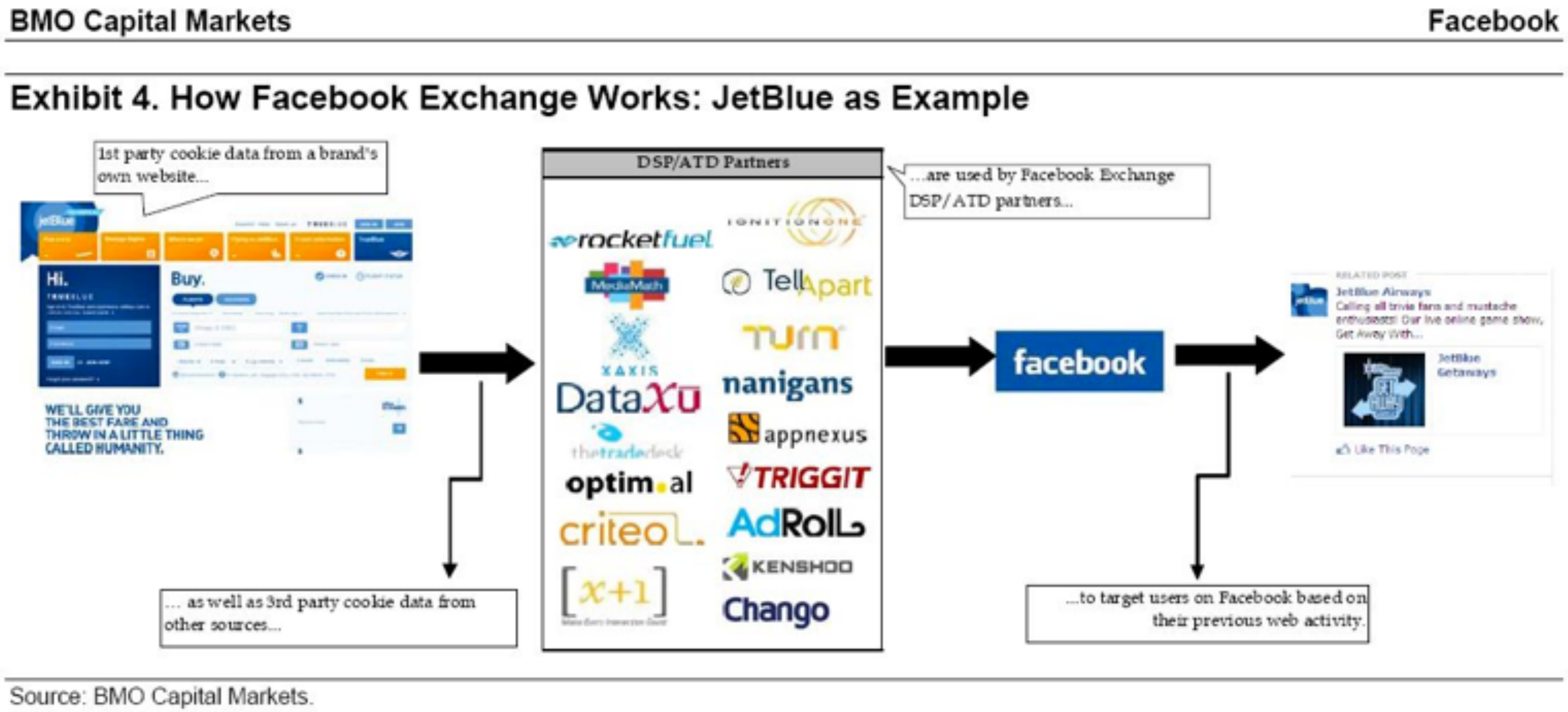
Thank you!

Questions?

Nathaniel Fruchter <fruchter@cmu.edu>
Hsin Miao <hsinm@andrew.cmu.edu>
Scott Stevenson <sbsteven@andrew.cmu.edu>
Rebecca Balebako <balebako@rand.org>

extra

Technical challenges



BMO Capital Markets

Facebook

Exhibit 4. How Facebook Exchange Works: JetBlue as Example

1st party cookie data from a brand's own website...

... as well as 3rd party cookie data from other sources...

...are used by Facebook Exchange DSP/ATD partners...

...to target users on Facebook based on their previous web activity.

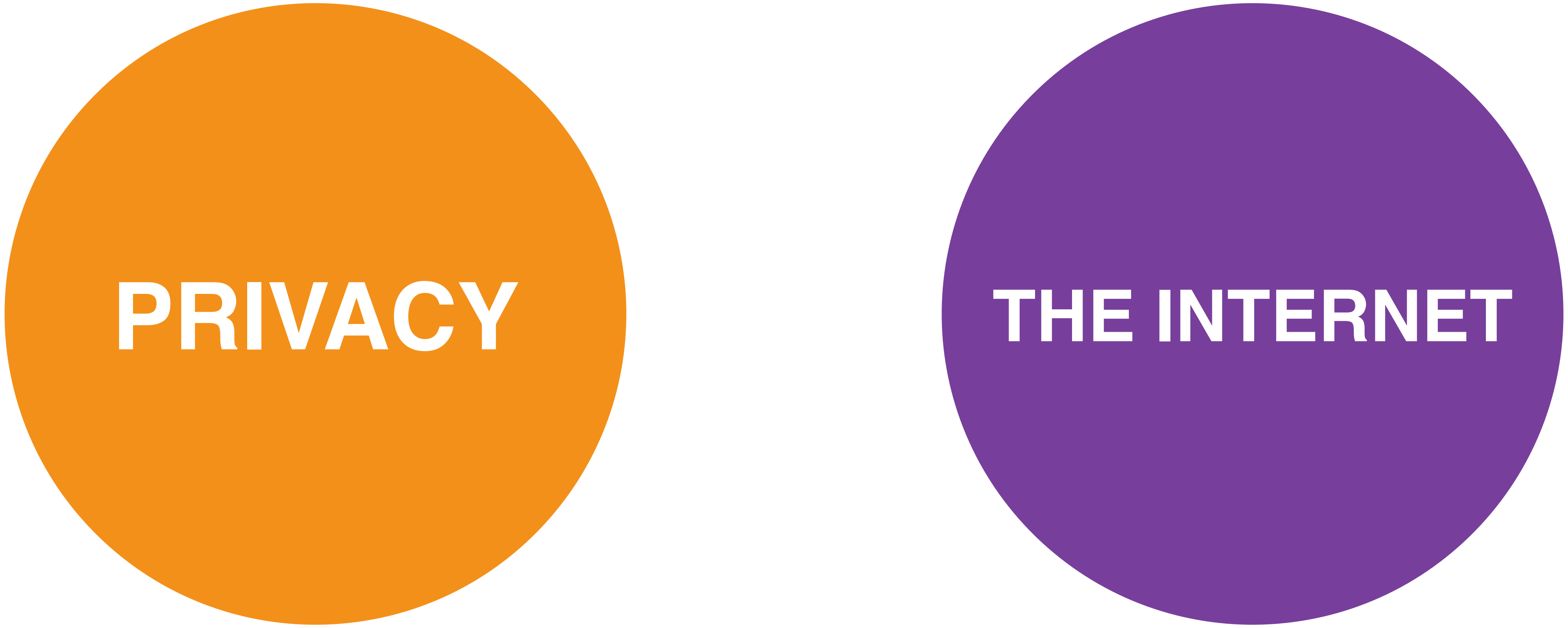
Source: BMO Capital Markets.

http://www.businessinsider.com.au/how-facebooks-fbx-ad-exchange-works-2013-1

Our approach

Network infrastructure

- How do you make it look like your connection is coming from a certain country?
- Tor is a possibility, but messy to work with
  - Uncertainty at endpoints with exit nodes
  - Connection can be slow or intermittent
- Sourcing VPNs raises other issues
  - Can interfere with traffic, cost money



PRIVACY

THE INTERNET

AN OPTIMISTIC VENN DIAGRAM

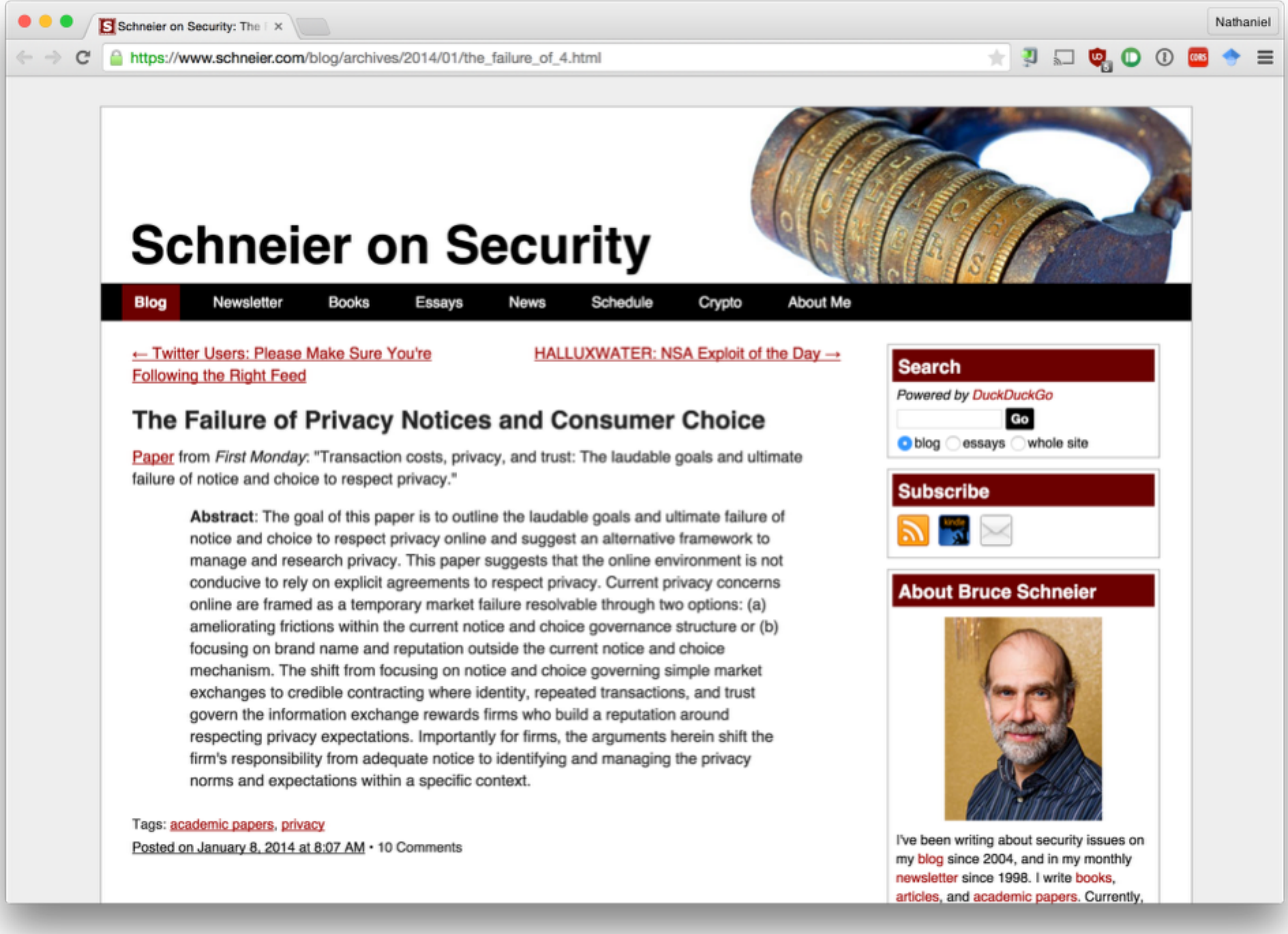
“Privacy is a value so complex, entangled in competing and contradictory dimensions, so engorged with various and distinct meanings… that I sometimes despair whether it can be usefully addressed at all.”

—Robert C. Post

Three Concepts of Privacy, 89 GEO. L.J. 2087, 2087 (2001).

Technical challenges

- **Is online / web activity deterministic?**
- Page loads
- People
- Devices
- Locations
- Internet connections
- The list goes on…


Schneier on Security: The
Nathaniel
https://www.schneier.com/blog/archives/2014/01/the_failure_of_4.html

Schneier on Security

Blog
Newsletter
Books
Essays
News
Schedule
Crypto
About Me

← Twitter Users: Please Make Sure You're Following the Right Feed

HALLUXWATER: NSA Exploit of the Day →

The Failure of Privacy Notices and Consumer Choice

Paper from *First Monday*: "Transaction costs, privacy, and trust: The laudable goals and ultimate failure of notice and choice to respect privacy."

> **Abstract**: The goal of this paper is to outline the laudable goals and ultimate failure of notice and choice to respect privacy online and suggest an alternative framework to manage and research privacy. This paper suggests that the online environment is not conducive to rely on explicit agreements to respect privacy. Current privacy concerns online are framed as a temporary market failure resolvable through two options: (a) ameliorating frictions within the current notice and choice governance structure or (b) focusing on brand name and reputation outside the current notice and choice mechanism. The shift from focusing on notice and choice governing simple market exchanges to credible contracting where identity, repeated transactions, and trust govern the information exchange rewards firms who build a reputation around respecting privacy expectations. Importantly for firms, the arguments herein shift the firm's responsibility from adequate notice to identifying and managing the privacy norms and expectations within a specific context.

Tags: academic papers, privacy

Posted on January 8, 2014 at 8:07 AM • 10 Comments

Search

Powered by DuckDuckGo

Go

● blog ○ essays ○ whole site

Subscribe

About Bruce Schneier

I've been writing about security issues on my blog since 2004, and in my monthly newsletter since 1998. I write books, articles, and academic papers. Currently,

https://www.schneier.com/blog/archives/2014/01/the_failure_of_4.html

Next steps

- How does culture affect Internet use?
- How do we intersect this with businesses' data collection habits?